REMARKS

As noted previously, the Applicant appreciates the Examiner's thorough examination of the subject application.

Claims 1, 5-13, 26, 28, 29 and 33 remain in the application. In the final Office Action mailed 17 March 2008, claims 1, 5-13, 26, 28, 29, and 33, were rejected on statutory grounds, as described in further detail below. Claims 1, 26, and 29 are amended herein. No new matter has been added.

Applicant respectfully requests reconsideration and further examination of the application based on the foregoing amendments and the following remarks.

Claim Rejections – 35 U.S.C. § 103

Concerning items 1-2 of the Office Action, claims 1, 5-13, 26, 28, 29, and 33 were rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 6,480,598 to Reding et al. ("Reding") in view of U.S. Patent No. 6,725,194 to Bartosik et al. ("Bartosik"). Applicant respectfully traverses this rejection and requests reconsideration for the following reasons.

For a rejection under 35 U.S.C. § 103(a), the cited reference(s) must teach or suggest each and every limitation of the claim(s) at issue and proper motivation must exist to combine or modify the teachings of the references in the way proposed for the rejection. Stated another way, a conclusion of obviousness requires that the reference(s) relied upon be enabling in that it/they put the public in possession of the claimed invention and proper motivation must exist to combine or modify the teachings of the references in the way proposed by the Examiner. These requirements are not met in this situation, as will be explained.

i. Claimed Subject Matter

To clarify inventive aspects of the claimed invention, the independent claims have been amended to include the limitation that the speech recognition functionality is "speaker-independent". Applicant notes that such limitations were previously introduced to the claims by the Amendment of 23 November 2005 and that the limitations were withdrawn because the Examiner construed such to

- 7 -

be new matter.

Applicant asserts, however, that "speaker-independent" is not new matter, and at the very least the subject application as filed inherently supports such a limitation. For example, there are several/many references in the specification of the subject application to the systems/methods applying to use with SR **over the telephone**, **connections by telephone**, **phone calls**, etc. By definition, SR used on a phone call **must be speaker independent** – because the system does not control (and can't know) who will answer a particular phone call (in fact in many calls the person answering hands the phone off to someone else, so the system has to SR more than one person in the same call). This is as compared to a speaker-dependent SR where the same known person speaks into a PC, as is done with dictation.

ii. Cited Art

The Reding reference is directed to a system/method for handing in-bound ("IB") calls to have an automated system handle some/all the calls that live operators might handle, during different parts of the day. The automated system of Reding mimics the connection protocol of the live agent's login. One can adjust the mix of live agents vs. automated agents on a given network, etc. In fact, Reding talks about transferring automated calls that are not going well back to a live agent so they can complete the call – rather than use the data/recording to improve the SR, they just pass the live call itself off to human. Reding appears to be all about customer service in the moment (use agents or SR, if SR is not working transfer to an agent) - not improving SR off-line like as does Applicant's claimed invention. Reding does not teach or suggest diverting recorded utterances based on confidence factors, let alone using such data to improve SR accuracy.

In contrast with Applicant's claims, **Bartosik teaches a speaker-dependent dictation system**, that is, a method/system for speaker-dependent speech recognition ("SR") that the SR user operates by speaking into a microphone attached directly to a computer. The method/system of Bartosik is for a way to correct dictated text, by having a human operator view what was processed by the speaker-dependent SR and type in corrections. Bartosik also describes a method for using the

corrected text to optimize the recognizer for a specific user's voice. It also presents ALL the text that the user spoke (as it is processing a continuous stream of dictated words) and not specific choices from short utterances (flagged text) as recited in Applicant's claims. The systems/apparatus taught by Bartosik are in stark contrast with Applicant's claimed systems/methods, which are **speaker-independent** and which are useful for receiving calls, i.e., when there is no control over who is calling in.

While Bartosik does not expressly describe its techniques as employing "speaker-dependent" SR, one skilled in the art would clearly understand the reference to use such. For reference to dictation and speaker-dependent SR, see Bartosik, for example, at the following locations:

(1) FIG. 1, elements 2, 42, and 54 showing a microphone connected to a computer:

(2) Col. 1, lines 13-16, describing a microphone and line 56 reciting "adjusting the speaker";

(3) Col. 3, lines 10, 12, 18, 41, 47, and 55 ("dictation . . .", "microphone", and "USB" for connecting the microphone to a computer);

(4) Col. 6, lines 47-54, where the user is said to train the system based on the user's voice;

(5) Col. 8, line 63, describing "dictations";

(6) Col. 12, line 45, stating that the Bartosik system "forms a dictating machine";

(7) Col. 13, lines 10-12 stating that the Bartosik system is "adjusted to the respective user"; and

(8) Col. 14, lines 3-5, stating that the microphone for the system is connection via USB to the computer.

iii. All claim elements are not taught/suggested

Amended claim 1, representative of the independent claims of the application, recites:

A speech recognition system comprising:

a querying device for posing at least one query over a telephone to a telephone respondent;

a speech recognition device that is configured and arranged to receive an audio response from said respondent over the telephone and to conduct a **speaker-independent speech recognition analysis** of said audio response to automatically produce a corresponding text response;

a storage device for recording and storing said audio response as it is received by said speech recognition device;

an accuracy determination device for automatically comparing said text response to a text set of expected responses and determining whether said text response corresponds to one of said expected responses, **wherein said accuracy determination device is configured and arranged to determine whether said text response corresponds to one of said expected responses within a predetermined accuracy confidence parameter and to automatically flag said audio response so as to produce a flagged audio response for further review by a human operator, wherein the human operator is different from the telephone respondent, when said text response does not correspond to one of said expected responses within said predetermined accuracy confidence parameter**; and

a human interface device for enabling said human operator to hear said flagged audio response and review the corresponding text response for the flagged audio response to determine the actual text response for the flagged audio response, either by selecting from a pre-determined list of text responses or typing the actual text response if no such match exists in the pre-determined list of text responses.

[Emphasis added]

For the rejection, the Examiner cites the Reding as the primary reference. As described previously, Reding is directed to methods and apparatus providing operator (as in human operator) services to callers in a fully or partially automated manner. As described above, Reding fails to teach or suggest the use of confidence factors as claimed by the Applicant, e.g., in amended claims 1, 26, and 29. The Examiner implicitly admits that Reding fails to teach the confidence parameter recited in Applicant's independent claims.

In an attempt to ford this gap in the teachings of Reding, the Examiner alleges that Bartosik teaches, *inter alia*, "an accuracy determination device for automatically comparing said text responses to a set of text responses . . .within a predetermined accuracy confidence parameter," citing Bartosik at col. 6, lines 7-16 and col. 9, lines 1-62. Applicant respectfully traverses such a characterization of Bartosik.

Applicant notes that Bartosik does not teach the above-quoted portion of amended claim 1 but rather is seen as **actually teaching systems and methods that functions similar to a dictation machine**. *See, e.g.*, Bartosik, col. 3, lines 8-11 ("FIG. 1 shows a computer 1 by which a speech recognition program according to a speech recognition method is run, which computer 1 forms a dictating machine with a secondary speech recognition device."). Bartosik relies upon a user reading all recognized text information to determine erroneous recognitions, and because of such actually teaches away from the Applicant's claims.

Moreover, Applicant contends that attributing a flag to a portion of speech not matching a response from a set of anticipated responses does not read on or correlate to attributing a sliding scale factor to such portion of speech as taught by Bartosik. As noted above, Applicant's independent claims, e.g., claim 1, recite that "the human operator is different from the telephone respondent."

Thus neither Reding nor Bartosik teach or suggest all of the limitations of claims 1, 5-13, 26, 28, 29, and 33 of the subject application.

iv. Proper motivation is not present

Not only do Reding and Bartosik fail teach or suggest all of the limitations of Applicant's claims, but proper motivation does not exist to combine and/or modify the teachings of the references in the way the Examiner has proposed.

While acknowledging deficiencies of Bartosik, the Examiner provided the following as ostensible motivation for the obviousness rejection:

Therefore, it would have been obvious to one of ordinary skill in the art to

- 11 -

incorporate the automatic flagging feature of Bartosik into the recognition process of Reding because it would advantageously improve the recognition of the device . . . as well as improving the automation of the recognition process, which is a concern in Reding et al.

Applicant respectfully submits that the Examiner's logic employed here is conclusory at best and indicative of impermissible hindsight analysis at worst.

Bartosik does not teach the flagging and confidence parameter limitations of Applicant's claims, and actually makes clear that it is fundamentally different relative to Applicant's claimed invention: namely, that the systems and methods of Bartosik derive a numerical value (the correspondence indicator CI) that is used for the adjustment of a speech coefficient indicator SKI during operation in a training mode – this correspondence indicator (CI) is not used to flag an audio response in the way claimed by Applicant:

> Furthermore, the text comparing means 52, when comparing the recognized text information RTI and the corrected text information CTI, **are provided for determining a correspondence indicator CI for each text part.** The text comparing means 52 then determine how many matching words featured by a grey field a text part contains. Furthermore, the text comparing means 52 determine penalty points for each text part, with one penalty point being awarded for each insertion, deletion or substitution of a word in the corrected text information CTI. **The correspondence indicator CI of the text part is determined from the number of the corresponding words and penalty points of a text part.**
>
> In the text comparing means 52 is determined a minimum value MW for the correspondence indicator CI, which minimum value is fallen short of when for a text part more than three penalty points are awarded for corrections of adjacent words of the corrected text information CTI. **For the adjustment of the speech coefficient indicator SKI, only text parts are used whose correspondence indicator CI exceeds the minimum value MW.**

(Bartosik, col. 9, lines 43-62) [Emphasis added]

Bartosik further explains that the adjustment of the SKI occurs in a training mode – not a normal use mode:

> **When the initial training mode is activated,** the text processing means 47

- 12 -

are arranged for reading out the training text information TTI from the training-text memory means 47 and for feeding respective picture information PI to the monitor 4. A user can then utter the training text displayed on the monitor 4 into the microphone 6 to adjust the speech recognition device to the user's type of pronounciation [sic].

The speech recognition device has adjusting means 50 for adjusting the speech coefficient indicator SKI stored in the speech-coefficient memory means 38 to the type of pronounciation [sic]of the user and also to words and word sequences commonly used by the user. The text memory means 43, the correction means 49 and the adjusting means 50 together form the training means 51. Such an adjustment of the speech coefficient indicator SKI takes place when the initial training mode is activated in which the training text information TTI read by the user is known.

Such an adjustment, however, also takes place in an adjustment mode in which text information corresponding to voice information is recognized as recognized text information RTI and is corrected by the user into corrected text information CTI. **For this purpose, the training means 51 include text comparing means 52, which are arranged for comparing the recognized text information RTI with the corrected text information CTI and for determining at least a correspondence indicator CI.** In the text comparing means 52 an adjustment table 53 shown in FIG. 4 is established when the adjustment mode is on, which table will be further explained hereinafter.

(Bartosik, col. 6, line 47 through col. 7, line 9.) [Emphasis added]

Thus, Bartosik teaches away from Applicant's amended claims, e.g., for at least the limitations of "speaker-independent" and the flagging and confidence parameter limitations of Applicant's claims.

Applicant reasserts that proper motivation is not present when one or more of the references teach away from the structure/modification suggested by the Examiner, as is the present case concerning Bartosik and Papineni. MPEP § 2145 (X)(D)(2) explains "**It is improper to combine references where the references teach away from their combination,**" citing *In re Grasselli*, 713 F.2d 731, 743, 218 USPQ 769, 779 (Fed. Circ. 1983), which case was decided after and is consonant with *In re Keller*, 642 F.2d 413, 208 USPQ 871 (CCPA 1981).

For at least the foregoing reasons, the cited combination of Reding and Bartosik (regardless of whether the references are considered together or separately) is an improper basis for a rejection of

claims 1, 5-13, 26, 28, 29, and 33 under 35 U.S.C. § 103(a); Applicant requests that the rejection of these claims be removed accordingly.

Conclusion

In view of the remarks and amendments submitted herein, Applicant respectfully submits that all of the claims now pending in the subject application are in condition for allowance, and therefore requests a Notice of Allowance for the application.

Authorization is hereby given to charge any required fees and to credit any overpayments to deposit account No. 50-1133. If the Examiner believes there are any outstanding issues to be resolved with respect to the above-identified application, the Examiner is invited to telephone the undersigned at his earliest convenience so that such issues may be resolved.

Respectfully submitted,

McDERMOTT WILL & EMERY LLP

Date: 20 July 2009
　　　　　　　　　　　　　　　　　/G. Matthew McCloskey/_____
　　　　　　　　　　　　　　　　　Toby H. Kusmer, P.C., Reg. No. 26,418
　　　　　　　　　　　　　　　　　G. Matthew McCloskey, Reg. No. 47,025
　　　　　　　　　　　　　　　　　28 State Street
　　　　　　　　　　　　　　　　　Boston, MA 02109
　　　　　　　　　　　　　　　　　V: (617) 535-4082
　　　　　　　　　　　　　　　　　F: (617) 535-3800